THE SCHOOL REVIEW

A JOURNAL OF SECONDARY EDUCATION

VOLUME XX
NUMBER 5

MAY, 1912

WHOLE
NUMBER 195

THE MEASUREMENT OF EDUCATIONAL PRODUCTS[1]

EDWARD L. THORNDIKE
Teachers College, Columbia University

The business of man, I suppose, is to change the world for the better. Things and men have to be reformed so that they will satisfy better those wants which reason finds worthy. And man's wants also must be changed. The guide of reason in changing things to fit human wants, and in changing men themselves to fit their nobler possibilities and the common good, is science—matter-of-fact, verifiable knowledge of the nature and causes of changes in things and men.

Education, like history, economics, sociology, and the other sciences of man, is just beginning to give promise of quantitative knowledge, of descriptions of facts as numerically defined amounts, and of relations or laws in terms of rigid, unambiguous equations. The changes that take place in intellect and character are coming to be measured with the same general technique, and, we may hope, with the same passion for clearness and precision, which has served the physical sciences for the last two hundred years.

I shall try today to describe one feature of this quantitative work in the science of education—perhaps I should say, this quantitative work toward a science of education. I wish that my description might be adequate, so that those of you who esteem the definiteness and rigor of the physical sciences would be led to agree

[1] Paper read at the meeting of the Harvard Teachers' Association, Cambridge, Massachusetts, March 9, 1912.

with me that a science of the changes produced in man by education may be made just as definite and exact—and so that those of you who esteem the significant and universal application of the humanities would admit that the life of man gains a new dignity when the insights of the dramatist, sage, or statesman are reinforced by measuring-rods and scales for weighing and accounting for human nature and behavior.

The need for measurement of the differences made by education —of educational products—is equally demonstrable whether our concern is with education as a science or education as a business. However it is defined, education concerns the production and prevention of changes in human beings; and a science of education must identify these changes, compare them, and relate them to their causes. To do this it must measure them. The balance-sheet of education as a business has as its credit side certain educational products—knowledge of such and such school subjects, habits of courage or persistence, and the like, skill in this or that trade or profession, added interests, refinements of taste, ideals of honor, service, and truth. It is fruitless to keep only the debit account—of time and money expended, of teachers, books, supplies, buildings, courses of study, and methods of teaching and the like, leaving the credit account—the results achieved, the products of education—vague and insecure.

There are peculiar difficulties in keeping an accurate account—in measuring the changes which are the data for the science of education. The facts are extraordinarily complex, very widely variable, and do not at all readily suggest units, scales, or graded standards by means of which they may be identified, compared, and related.

So apparently simple an ability as ordinary addition of integers can be shown to require analysis into at least nine separate abilities, each of which probably requires further analysis, in one case, into perhaps ninety component ability-atoms. The achievement of any pupil in any matter of intellect, character, or skill usually varies widely, so that many measurements of ostensibly the same fact have to be made to relieve it from its chance or accidental error. Scales, graded standards, by which to report knowledge of German,

ability to spell, skill in cooking, original power in mathematics, appreciation of music, or any educational fact you may think of, are now where the thermometer, spectroscope, and galvanometer were three hundred years ago—they do not exist.

The beginnings that are being made toward overcoming this third difficulty are my special topic today. I ask you to consider the requirements of a valid scale of measurement in general, and to consider whether we can meet them in measuring educational products.

The ordinary scale for weight, zero, one gram, two grams, three grams, and so on, exemplifies an ideal scale in four respects: First, it is a series of perfectly definable facts. All men over all the world may know exactly what is meant by two grams or four grams. In the second place, each amount is a different amount of the same kind of thing. "Four grams" is so much of the same kind of thing of which "two grams" is "two grams." In the third place, the differences between any two of the amounts are perfectly defined in terms of some unit of difference. The step from four to five grams is the same as the step from forty-seven to forty-eight grams, and so on. Lastly, the zero point of the scale is absolute; that is, it means just barely not any of the thing in question. I shall take these points in order.

Can we, then, in a science of education, get a series of perfectly defined points of the amount of some thing, so that all men may know what each man means by the statement he makes, as all know it in the case of "one gram" or "two grams"? At the present time we have certain descriptions or identifications of these points on the scale, or amounts of the thing. We say, in the case of handwriting, "a good plain hand," which is about as useful a statement—of about the same character—as if, in describing the weight of a body, we should say, "as heavy as a baby." In describing the ability of a pupil in German, we say "His knowledge is about equivalent to that required for passing intermediate German." That is about as if, in describing the length of an object, we should say, "It is about as long as a man can jump." In measuring a Freshman's composition in English in Harvard College, we say it is of "C" grade. That is about on a level with the statement that

a glass of water is "tepid;" and although there may be a certain emotional appropriateness in the use of the word "tepid" to describe the Freshman's composition, as a scientific measure it is hardly adequate.

Now it is an easy task, theoretically, for educational science to take these vague, ambiguous statements of common-sense and refine them as physical science has in the past refined similar measures in the case of physical facts. We can, for example, define a good plain hand by printing a sample of it. If I should write on the blackboard with a certain degree of legibility and beauty, and if this association were to say, "That is what we call a good plain hand"; and if we were to have copies made of it, so that people could see it; and if we should get the people of the United States and of other countries to agree that "a good plain hand" meant a hand as good as, and no better than, the printed sample in question —then we would have one point on the scale thereby defined.

In the case of spelling, we can define a point on the scale as the ability to spell words as hard as, but no harder than, "at" and "go," or "wish" and "touch," and so on to "millinery," "development," or words of any difficulty we choose. We could, in the case of German, agree upon a series of passages, graded in difficulty of translation, and say, "What we mean by ability 'four' is the ability to translate *this* passage with a certain degree of precision, but not to translate the next more difficult passage," and so on. Such a series of passages would be easy enough to get, and would define for us points in this particular scale. Similarly we could get a series of originals in mathematics, graded in difficulty, which would define a series of points on the scale of mathematical ability. We could in the same way define the amount of merit in an English composition by an actual sample. Lately my friend Mr. Hillegas has been engaged in doing that in the case of English writing by young people in their teens.

The second requirement of the ideal scale, you will remember, was that the different amounts should be amounts of the same thing. Here again the present practice is far from as advanced as we would wish. Take the case of ability in arithmetic, in grammar-school pupils, for example. As you well know, our measurement of

ability in arithmetic actually is a measurement of two different things: sheer mathematical insight and knowledge, on the one hand; and acquaintance with language, on the other. For example: suppose I give this problem to pupils in the second grade at school, "If you already had two cents, and someone gave you three cents too, what sum would you then have in all?" Suppose that a day later, I give the problem, "How many are three and two?" As you all know, the percentage of children getting the second example correct would be far higher than in the case of the first example. The first test is not so much in arithmetic as in language; the difficulty is in the words "already," "too," "someone," "would," "in all."

In the case of spelling, ordinary measurements of achievement mix up knowledge of spelling proper with an acquaintance with words and meanings. For example, to be able to spell the words "too" and "there" requires mainly ability in spelling; but the ability to spell correctly the words "Popocatapetl," "Abracadabra," would require to some slight degree a knowledge of spelling, but largely a wide knowledge of the language.

Those of you who teach languages in secondary schools are well aware that the ordinary college-entrance examination and also your own tests of achievement in a language are scales which measure two things. Part of the difficulties which pertain to translating a passage in a foreign language arise from the difficulty of understanding the general thought of the passage. We have all, I fear, given our pupils passages which, so to speak, they could not translate even if they were given in English! They would not know what they meant, if they were given in English. The difficulty would be that they did not have the general knowledge of affairs and things to understand the meaning of the passage in any tongue. Here again there is nothing impossible in the task. We can separate out, by skilful experiments, achievement in mastering spelling difficulties proper, apart from the difficulties of mastering a general acquaintance with English words. We can devise passages such that the facts and relations are simple, but the German, the French, or the Latin, is difficult; and other passages where the French or German or Latin is easy, but the meaning is

difficult. We can separate into two scales these two sets of facts measured.

For the purpose of what I have to say, let us assume, then, that we are measuring the same sort of thing, and that we have reached a point where we have, as our identified quantities, or amounts, or "scale-points," things which are in a linear series, and which differ *in amount only*, being of the same quality or thing or product.

The third requirement of the ideal scale was that the differences between any two scale-points or values should be rigorously and perfectly defined in terms of some unit of difference. Now we come to a point where very little has been done by our educational practice. Most of the quantities that we assign in education signify only relative positions, really. We perhaps make a pretense at their being differences in amount. When we say that one pupil is "75" and another is "80," a third is "85," another is "90," and another "95," if there is one thing we do know, it is that the step from 98 to 99 is, as a matter of fact, almost never equal to the step from 59 to 60, or 49 to 50. We do not know what their proper ratios are, however. Nobody here could tell from any given set of grades or averages of grades what the exact relation of the difference between 65 and 70 was to the difference between 85 and 90. We do not know, with respect to general educational products, whether the step from grammar-school to high-school graduation is equal to, less than, or greater than, the step from high-school graduation to college graduation; much less can we put the two steps into an exact ratio. We do not know whether the step from a barely legible handwriting up to a good plain hand is equal to, less than, or greater than, the step from a good plain hand to a perfect, copper-plate writing; much less could we put the differences in terms of an exact ratio.

The problem, then, is to take the differences in relative merit that we all observe in school products and turn these into precise differences of amount. This can be done. Here, I shall have to take you into a somewhat technical discussion for a few moments. In fact, to deal with the subject properly, I should have to go through a rather laborious discussion of the logic of quantity generally. I take equality of steps of difference, which is the simplest case

rhetorically, though all I say applies to getting any ratios between differences defined. When a difference is not always noted by competent judges, we may say that two differences are equal which are equally often noted. If, for example, we had four samples, 1, 2, 3, and 4, of English writing, such that 80 per cent of an expert group said that 1 was better than 2, whereas 20 per cent said that 2 was better than 1, and 80 per cent said that 3 was better than 4, and 20 per cent that 4 was better than 3; then, in a very important, and, with certain limitations, true sense, we can say that the difference between 1 and 2 is equal to the difference between 3 and 4, because it is equally often noticed by equally competent judges. This method of turning measurements by relative positions into measurements in terms of units of amount was first used, very roughly, by the late Mr. Francis Galton, an eminent English man of science. It has been used by Professor Cattell and others at Columbia University; and it is used in many fields of human life, apart from education. It is the most convenient tool whereby to express the steps of difference, and define the distances of one point from another, of all the educational scales that we have. I shall illustrate this by reading six compositions by pupils of high-school age, which, by such use of the opinion of two hundred experts—common-sense people, experts in education, teachers, literary men, editors, and the like—are found to differ by approximately equal steps. These steps are not absolutely equal, being in fact 11, 10, 11, 9, and 10. I will read the compositions, in their order, from the lowest up:

26. "Advantage evils are things of tyranny and there are many advantage evils. One thing is that when they oppress the people they suffer awful I think it is a terrible thing when they say that you can be hanged down or trodden down without mercy and the tyranny does what they want there was tyrans in the revolutionary war and so they throwed off the yok."

37. "When Sulla came back from his conquest Marius had put himself consul so sulla with the army he had with him in his conquest siezed the government from Marius and put himself in consul and had a list of his enemys printy and the men whoes names were on this list we beheaded."

47. "First: De Quincys mother was a beautiful women and through her De Quincy in hereted much of his genius.

"His running away from school enfluenced him much as he roamed through the woods, valleys and his mind became very meditative.

"The greatest enfluence of De Quincy's life was the opium habit. If it was not for this habit it is doubtful whether we would now be reading his writings.

"His companions during his college course and even before that time were great enfluences. The surroundings of De Quincy were influences. Not only De Quincy's habit of opium but other habits which were peculiar to his life.

"His marriage to the woman which he did not especially care for.

"The many well educated and noteworthy friends of De Quincy."

58. "The passages given show the following characteristic of Fluellen: his inclination to brag, his professed knowledge of History, his complaining character, his great patriotism, pride of his leader, admired honesty, revengeful, love of fun and punishment of those who deserve it."

67. "Ichabod Crane was a schoolmaster in a place called Sleepy Hollow. He was tall and slim with broad shoulders, long arms that dangled far below his coat sleeves. His feet looked as if they might easily have been used for shovels. His nose was long and his entire frame was most loosely hung to-gether."

77. "When at last it did arrive the postmaster began to quickly sort the bundles, we waited anxiously. Immediately upon receiving our bundles, I lashed the horses and they responded with a jump. Out into the country we drove at reckless speed—everywhere spreading like wildfire the news, 'Victory.' The exileration that we all felt was shared by the horses. Up and down grade and over bridges, we drove at breakneck speed and spreading the news at every hamlet with that one cry 'Victory.' When at last we were back home again, it was with the hope that we should have another ride some day with 'Victory.'"

The second sort of argument is somewhat more intricate. Suppose that we have a scale running from a low to a high degree of achievement, say, in solving geometrical problems. And suppose that we know the geometrical form of the distribution of, say, fourth-year high-school pupils with respect to that achievement. Let this be a rectangle, a half circle on the linear scale as base, the normal probability-surface, or any other. Then, from the percentages of such pupils solving problems A, B, C, D, etc., we can determine the ratios $\frac{A-B}{B-C}, \frac{B-C}{C-D}, \frac{C-D}{D-E}$, etc.

Thus suppose the form of the distribution of ability to be that defined by the scale line as a base and the line $Y=Pe^{\frac{-x^2}{2npq}}$, and call L the achievement of solving a problem which 9,987 out of 10,000 such pupils can solve. Call H the achievement of solving a problem such as only 13 out of 10,000 such pupils can solve.

To get a series of, say, 13 achievements varying from *H* to *L* by 12 equal steps we shall have to find problems which can be solved by 13, 62, 227, 668, 1,586, 3,085, 5,000, 6,915, 8,414, 9,332, 9,773, 9,938, and 9,987 out of 10,000 of the group respectively.

This method has so far been used only once or twice, but is likely to become of great importance in the case of certain educational products.

Neither of these methods of defining the difference between facts is that commonly used in the physical sciences. To the physicist, those differences are equal which are products of the same cause in the same circumstances or which, under the same conditions, produce the same effect. The difference between nine inches and ten inches is equal to the differences between six inches and seven inches because one standard length produces either difference. One degree of temperature is equal to any other degree of temperature, in the sense that it will raise the mercury the same distance. We can do something like this in education, though it becomes rather difficult, and careful logic is undoubtedly required.

The last element of a valid scale was that it should be referable to a zero which meant just barely not any of the thing in question —that is, an absolute zero. The importance of this may not appeal to many of you; it does not appeal to people generally. What an absolute zero does is to permit us to use the "times" judgment—to say that this educational achievement is twice or one-half or three-fourths times that; speaking algebraically, it enables us to use equations with two instead of four unknown quantities, because the zero points are true zero points.

If you ask anyone to tell you what he means, or show you what he means, by "just barely not any originality in mathematics," or "just not any merit in handwriting," he is likely to be amused, and say that he never thought of it, and does not intend to! But I must confess that to me it appears that such a person has not clearly defined his ideas. A person cannot think accurately or wisely about any quantity until he has referred it to an absolute zero on some scale. Now the fact that zero points do not stare us in the face in the case of mathematical originality, or knowledge of German, or ability in writing, as they do in the case of measures of

length, weight, and time, is no excuse for not trying to get them. They can be got, at least approximately. For example, if we define the zero of ability in spelling as the ability to spell a word like "go" or "so," we shall not be far astray, for if a child has reached a point where he can be measured at all in respect to spelling and is not able to spell these words, he is approximately at just not any ability at all in spelling. I think if we should agree that for a pupil just not to know the meaning of "ja" and "nein" under ordinary conditions of life in New England today would indicate that he was approximately at the zero point in knowledge of German, we should not be far astray. If he knew that "ja" meant "yes" but did not know that "nein" meant "no," he would be so close to the zero point that our error would be slight.

In handwriting, a handwriting which is recognizable *as* handwriting, but is of just not any beauty or legibility, is not hard to define. Fig. 1 may serve roughly. We should say that this was handwriting; it is not a house, nor a mere test of motor skill; it is, in a sense, handwriting; but it has no legibility or beauty. The zero point in the case of composition was determined by the judgment of forty experts, divided about equally between men of

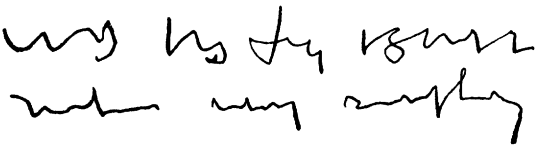

FIG. 1

affairs, expert psychologists, experts in education, practical schoolmen, and literary men themselves. The sample taken as zero of merit for writing by young people in their teens was:

Dear Sir: I write to say that it aint a square deal Schools is I say they is I went to a school. red and gree green and brown aint it hito bit I say he don't know his business not today nor yesterday and you know it and I want Jennie to get me out.

If we get scale points defined, and their distances defined, and establish an absolute zero, there is no further difficulty in constructing a scale for achievements of human nature. Such scales have every logical qualification that any of the scales for physical measurement have.

There is no limit, theoretically, to the kind of thing for which scales are practicable. I have chosen for convenience the simpler and easier cases. But the arguments apply equally to the sense of evidence in history, excellence of judgment in affairs, devotion to the common good, or any quality, no matter how complex, that one may take.

It may save the time of all of us if I say just a word or two about two or three objections.

It may be said, "All this is unnecessary; the good old adjectives are enough for educational work." All that I have to retort is that for the kind of educational work that the person who makes this objection usually wishes to do, probably the old adjectives *are* all that is necessary.

A stronger objection would be that the common-sense judgment of a first-rate man without these units and scales is better than the action of the stupid man or incompetent man, with them. That is, of course, true—that a good man can do better work without them than a stupid man with them. It is precisely the work of science to get good work done by those of us who are rather mediocre. Thanks to the progress of science, we can now solve problems that Aristotle could not solve. We should all prefer to have for our children a stupid doctor of today, who nevertheless understood the use of antiseptics and antitoxins, than Galen or Hippocrates, though in respect to common-sense there would be no choice.

The third objection is one that a certain type of person feels very keenly, namely, that the personal, spiritual work of education—the direct human influence that the pupil may get—is not in the domain of exact science. But that is not a valid objection. Mothers do not love their babies less who weigh them. We do not serve our country less faithfully because we take its census, survey its coast-line, or compute its resources. Education will of course always need its poets, its artists, its craftsmen, as well as its managers and men of science, but it needs these also. There is no reason why the artistic life should be impeded by the life of measurement. There is no reason why sculptors should sulk because surveyors prosper; or why a poet should feel badly because his odes are printed on paper whose length and breadth are known, and which is sold for a money price.